



Summative assessment versus formative assessment: An ecological study of physical education by analyzing state-anxiety and shot-put performance among French high school students

Maxime Mastagli, Didier Malini, Jean-Philippe Hainaut, Aurélie van Hoya,
Benoît Bolmont

► To cite this version:

Maxime Mastagli, Didier Malini, Jean-Philippe Hainaut, Aurélie van Hoya, Benoît Bolmont. Summative assessment versus formative assessment: An ecological study of physical education by analyzing state-anxiety and shot-put performance among French high school students. *Journal of Physical Education and Sport*, 2020, 20 (Supplement issue 3), pp.2220-2229 / Art 298. 10.7752/jpes.2020.s3298 . hal-02913350

HAL Id: hal-02913350

<https://hal.univ-lorraine.fr/hal-02913350>

Submitted on 8 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Original Article

Summative assessment versus formative assessment: An ecological study of physical education by analyzing state-anxiety and shot-put performance among French high school students.

MAXIME MASTAGLI¹, DIDIER MALINI², JEAN-PHILIPPE HAINAUT³, AURÉLIE VAN HOYE⁴,
BENOÎT BOLMONT⁵

^{1,3,5} Laboratoire Lorrain de Psychologie et de Neurosciences (2LPN), Université de Lorraine, FRANCE.

² Education Nationale, Académie de Nancy-Metz, FRANCE

⁴ APEMAC, Université de Lorraine, FRANCE

Published online: July 31, 2020

(Accepted for publication: July 22, 2020)

DOI:10.7752/jpes.2020.s3298

Abstract

Problem Statement: Previous works have questioned the function of assessing and its influence on students' perception of the assessment. Depending on whether the assessment is summative or formative, the impact on learning and emotions may be different. However, few studies have been conducted in a physical education context where learning could be both cognitive and motor. **Purpose:** The purpose of this pilot study is to compare the effects of summative and formative assessments on students' state-anxiety, and shot-put performance in physical education. **Method:** Twenty-one high school students in a formative assessment class ($M_{age} = 14.95$; $SD_{age} = .50$) and twenty-two in a summative assessment class ($M_{age} = 15.14$; $SD_{age} = .64$) participated in the study. The students' characteristics between the two classes were not statistically different. Students filled out questionnaires measuring trait-anxiety and state-anxiety. State-anxiety was measured at the beginning of the lesson. Performance was evaluated through one shot-put attempt at the end of the lesson. Data collection was carried out in three phases: during the first lesson (*i.e.*, beginning of the learning cycle), during the fifth lesson (*i.e.*, end of the learning cycle), and during a deferred evaluation (*i.e.*, four months later). Non-parametric tests were performed. **Results:** There was no difference in trait-anxiety between the two classes. At the beginning of the learning cycle, there was no difference in state-anxiety or performance. Results show that at the end of the learning cycle, both classes improved their performance. The formative assessment class felt less anxiety than the summative assessment class. Four months later, the summative assessment class's performance had improved, whereas the formative assessment class's performance had remained at the same level. Additionally, the summative assessment class's anxiety remained at the same level, whereas the formative assessment class's anxiety improved. **Conclusion:** These results can help physical education teachers to design more successful learning and emotional experiences during the learning cycle, focusing more on providing formative assessment. This paper also adds to the debate around formative good/summative bad, questioning the short- and medium-term benefits and drawbacks of formative and summative assessment.

Keywords: pedagogy; evaluation; grade; emotion; motor learning

Introduction

Previous works have questioned the function of assessing, “to measure or to learn?” (Broadfoot, 2002), and its influence on student's perception of the assessment (Harlen et al., 2002; Lau, 2016). Assessment relies on “a wide range of methods for evaluating pupil's performance and attainment including formal testing and examinations” (Gipps, 2002). In physical education (PE) as in other academic discipline, different types of assessment have been positioned on a continuum from “performance-based assessment” (*i.e.*, quality of judgment of the student's performance) to “informational assessment” (*i.e.*, specifying learning progress; Krijgsman et al., 2017). Along this continuum, two main types of assessment have been distinguished. Formative assessment which takes place during the course of teaching is essentially used to provide feedback in order to support the teaching/learning process (Alcalá et al., 2019; Gipps, 2002). Summative assessment, which takes place at the end of a course, is used to provide information on what students have learned and how well a course has worked (Gipps, 2002).

Depending on the type of evaluation, the impact on learning may be different. Indeed, learning strategies are less effective when students work to earn a degree rather than to learn (Harlen et al., 2002). Previous work has shown that students are more passive and less interested in the task during learning cycle that includes summative assessment (Benmansour, 1999). Summative assessment favors surface learning approaches and can harm students' learning quality (Knight, 2002). In contrast, formative assessment supports the learning

process (Black & Wiliam, 1998; Harlen et al., 2002; Kwon et al., 2017; Roderick & Engel, 2001). Formative assessment allows students to appreciate the standards that are expected of them (Yorke, 2003) and plays a powerful role in students' academic motivation and achievement (Cauley & McMillan, 2010). It has also been shown that formative evaluation could improve the speed of learning (Volante & Beckett, 2011). Performance evaluation could also provoke different emotional response (McDonald, 2001). Emotional response to summative assessment has been well documented. Students consider the lesson less enjoyable and valuable when they are exposed to summative assessment rather than a non-grading lesson (Krijgsman et al., 2017). In addition, students feel significant psychological stress in preparation for summative assessment (Chee-Cheong, 1999; Law, 2007). Thus, the majority of students approach summative assessment with negative feelings (Carless & Lam, 2014) such as fear (Krijgsman et al., 2017), uncertainty and psychic tension (Pollard & Triggs, 2001) or state-anxiety (Black & Wiliam, 1998; Harlen et al., 2002). In contrast, emotional response to formative assessment has been less documented (Leighton, 2019; Lentillon-Kaestner et al., 2018). To our knowledge, only Lentillon-Kaestner et al. (2018) show that students in PE classes without grades work less stressfully.

Research into the impact of assessment has been conducted among various populations: primary education (e.g., Pollard & Triggs, 2001; Putwain & Best, 2011), secondary school (e.g., Benmansour, 1999; Krijgsman et al., 2017; Reay & Wiliam, 1999), higher education (e.g., Alcalá et al., 2019). These works have also been conducted in various educational contexts such as mathematics (e.g., Benmansour, 1999; Putwain & Best, 2011) and PE (e.g., Krijgsman et al., 2017; Lentillon-Kaestner et al., 2018; Martos-Garcia et al., 2017). PE is a specific academic discipline that requires particular attention because students are prompted to practice sports, which present different cognitive and motor learnings. Among the recent studies in PE, Krijgsman et al. (2017) examined the influence of the performance grade on students' motivation and fear. Martos-Garcia et al. (2017) evaluated differences in students' perceptions of an educational innovation based on formative and peer assessment. Lentillon-Kaestner et al. (2018) analyzed the impact of various types of evaluation on several factors related to teaching and learning. In light of these latest studies and considering the benefits and the drawbacks of assessment, we identified an important unanswered question: how does the type of assessment influence specific motor learning and emotional states such as anxiety in the PE context?

The present study was a pilot study aiming to provide an insight into the influence of the grade in PE. It examined the potential effect of summative assessment and formative assessment on motor learning and emotional reaction. To this end, the study was conducted in an ecologically valid setting (*i.e.*, during authentic PE lessons) with shot-putting, requiring strength and closed motor skills (Knapp, 1963; Linthorne, 2001). The first aim was to observe the impact of the type of assessment on the level of shot-put performance and state-anxiety within both classes (one in formative assessment and one in summative assessment) at the end of a learning cycle. The differences between the two classes and their evolution were compared from the first lesson onward. Few studies on motor learning have measured the variation in performance between the learning period and a long-term retention period (Henz & Schöllhorn, 2016; Kok et al., 2020; Scholz et al., 2009). Thus, the second aim was to examine the influence of summative and formative assessment on shot-put performance and state-anxiety four months after the end of the learning cycle. Two hypotheses were put forward. First, the summative assessment class would be more anxious and would perform less well than the formative assessment class at the end of the learning cycle. Second, the summative assessment class would be more anxious and would perform less well than the formative assessment class four months later.

Material & methods

Ethical Considerations

The teacher gave informed consent to participate in the current study. In accordance with the Declaration of Helsinki (1975), all participants gave their full informed consent prior to the longitudinal study. All students could choose not to participate in the study. They could withdraw at any time without any consequences. All students gave their oral consent. Each student's legal guardian and the school head were informed and gave permission.

Participants

Two complete high school classes of twenty-four French students each were recruited due to their common characteristics: (1) same PE teacher, (2) similar schedule (8:00 am to 10:00 am and 10:10 am to 12:10 pm on the same day), (3) same sports activity (shot-put), (4) similar intervention program. All participants were men. In the first and second class, a formative and a summative assessment were employed respectively. As soon as a student was absent during the learning cycle, he was excluded from the study. Two students from the formative assessment class and one student from the summative assessment class were absent at least one lesson. Moreover, as soon as one student was identified as an outlier for any of the study variables, he was excluded from the study. Z-tests were performed and identified two outliers (*i.e.*, with a Z score of less than -3 or more than 3) in both classes (*i.e.*, one student each). The final study sample was composed of twenty-one students in the formative assessment class and twenty-two students in the summative assessment class. The students' characteristics were not statistically different between the two classes (see Table 1). Four students in the

summative assessment class and six students in the formative assessment class had experience of throwing disciplines in athletics in PE at middle school. None of the students had any other throwing experience.

Table 1. Mean (standard deviation) of students' characteristics

| | Formative assessment class | Summative assessment class | <i>p</i> |
|--------------------|----------------------------|----------------------------|----------|
| Age | 14.95 (.50) | 15.14 (.64) | .36 |
| Body height | 172.00 (7.40) | 174.23 (7.23) | .25 |
| Body weight | 62.60 (13.61) | 62.50 (8.33) | .71 |

*Note: age is expressed in years; height is expressed in centimeters; weight is expressed in kilograms *p* = *p*-value of Mann-Whitney *U* test*

Measures

Shot-put Performance. Shot-putting was performed with a four-kilogram shot-put. The throw was carried out without impetus, with both feet on the ground. The participants' performance was assessed using a weighted score, representing the technical quality of the throw without the bias of individual precision. The weighted score was calculated by subtracting the distance between the falling point and the limit line (i.e., technical quality of throw) from the distance between the falling point and an ideal axis that was perpendicular to the limit line (i.e., individual precision; Hubiche & Pradet, 1993).

Anxiety. The participants' trait-anxiety was assessed in the first lesson using the Y2 form of Spielberg's State-Trait Anxiety Inventory (STAI; Spielberg, 1983). Trait-anxiety represents how subjects feel generally and reflect their disposition to developing state-anxiety (e.g., "*I feel that difficulties are piling*"; Spielberg, 1983). The participants' state-anxiety was assessed using the Y1 form of the STAI. State-anxiety represents how subjects feel in response to actual or non-actual threats (e.g., "*I feel upset*"; Spielberg, 1983). These self-evaluation questionnaires comprised 20 items each with a scale where 1 was "almost never" and 4 was "almost always". Scores were transformed according to the French high school student's calibration (Bruchon-Schweitzer & Paulhan, 1993).

Procedure

The study was conducted during a shot-put learning cycle of five lessons and during a deferred evaluation that took place four months after the end of the learning cycle. The learning cycle and the deferred evaluation took place in a stable environment, sheltered from random variables: indoor, with an ambient temperature of 18 degrees Celsius. After dressing in sportswear, students were asked to fill out the questionnaire assessing their level of anxiety. The teacher was present but did not interact with the students. Then a generalized and specific warm-up were carried out. This warm-up included races, jumps, and throws, and was strictly the same in both classes. Next, a technical sequence was analyzed and the students from both classes learned how to integrate this technical sequence into the complete throw by practicing with the reference weight. In addition, during these technical sequences, students were also trained using a three- to five-kilogram shot-put. Other exercises were performed without movement repetition such as throwing without any implement or with a javelin and a medicine ball. These exercises followed the differential learning approach recommending fluctuations during the motor learning process (Henz & Schöllhorn, 2016), learning by repetition (Gentile, 1972), learning through a series of methodical exercises (Gaulhofer & Streicher, 1924) and variability of practice (Ivanovici & Weissbluth, 2016; Raiola, 2017; Schmidt, 1975). These recommendations were implemented equally in both classes. Finally, a single throw was carried out, corresponding to the student's performance. The performance was measured at the end of the lesson, except for the first lesson where the student's performance was measured after the warm-up and before the analytical work. In this way, students from both classes had achieved a shot-put performance without training in the first lesson.

Indications on performance evaluation were given to both classes in the second lesson. These indications aimed to influence the students' state-anxiety. In the first and second class, a formative assessment and a summative assessment were employed, respectively. Formative assessment involved an informational shot-put performance at the end of each lesson, starting from the second lesson. Specifically, students in the formative assessment class were clearly informed that no grade was to be given and that the shot-put performance was only used to inform them of their progress in comparison to the previous lesson. Summative assessment consisted in giving a grade based only on their shot-put performance at the end of each lesson, starting from the second lesson. Students in the summative assessment class were clearly informed that a grade was given in each lesson and that this grade was only based on their shot-put performance.

Considering that sports performance and emotions are multifactorial, the impact of the type of assessment should be modest. For this reason, lessons two, three and four were not observed because it was likely that the effects of time and class could only be observed after several lessons. In sum, the measures were made in three phases: during the first lesson (i.e., beginning of the learning cycle), during the fifth lesson (i.e., end of the learning cycle), and during a deferred evaluation. The deferred evaluation took place four months after the learning cycle and was done just like in the first lesson (i.e., after dressing in sportswear, students were asked to

complete the questionnaire assessing their level of anxiety and the students' performance was measured after the warm-up and without any analytical work). During this deferred evaluation, there was no grade for the formative assessment class whereas the summative assessment students received a grade.

Data Analysis

All statistical analyses were performed with SPSS (version 22). Given the small sample size, non-parametric tests were conducted. To test our hypothesis, Mann-Whitney U tests were performed to test the class effect on shot-put performance and state-anxiety between the two classes at each data collection. Friedman's ANOVA and Wilcoxon tests were used with each class to explore time differences and time*class differences in state-anxiety and shot-put performance. Different highlights marked our analyses. Lesson 1 to lesson 5 was analyzed because this period represented the learning cycle for both classes. Next, lesson 5 to the deferred evaluation was also analyzed in order to observe the expected benefits and drawbacks of the formative and summative assessments, a few months later. For all statistical tests, the significance level was set at $p < .05$.

Results

Preliminary Analysis

No missing data were observed. The Y1 and Y2 forms of the STAI had satisfactory internal consistency (Table 2; $\alpha > .70$; Nunnally, 1978). The group effect between students who had experience in throwing (*i.e.*, four students in the summative assessment class and six students in the formative assessment class) and the others was examined for shot-put performance and state anxiety at each data collection. The Mann-Whitney U tests revealed that there was no group effect between students having throwing experience and students with no throwing experience.

After the French calibration of the STAI-Y scores (Bruchon-Schweitzer & Paulhan, 1993) presented in Table 2, students' mean score in trait-anxiety was considered as "medium". The Mann-Whitney U test revealed no difference in trait-anxiety between the two classes. Therefore, this variable was not included in the multivariate analyses. The level of state-anxiety was also considered as "medium".

Class Effect for Each Variable

Mann-Whitney U tests were performed in order to analyze the differences in performance and anxiety between the two assessment conditions (see Figure 1 and Table 2). No significant difference was observed in the level of performance and state-anxiety between the two classes during the first lesson. In the last lesson of the learning cycle, no significant difference was observed in the level of performance, whereas there was a significant difference in state-anxiety ($U = 313.50$; $z = 2.02$; $\eta^2 = .09$; $p = .044$) between the two classes, with the formative class less anxious than the summative class. No significant difference was observed in the level of performance and state-anxiety between the two classes during the deferred evaluation (*i.e.*, four months later).

Table 2. Reliability of the STAI Y1 and Y2, mean (standard deviation) of variables and test-differences between the two classes

| | α | Formative assessment class | Summative assessment class | η^2 | p |
|----------------------------|----------|----------------------------|----------------------------|----------|------|
| <i>Lesson 1</i> | | | | | |
| Performance | - | 540.86 (178.09) | 547.23 (171.15) | - | .94 |
| State-anxiety | .92 | 46.71 (6.29) | 46.73 (5.84) | - | .94 |
| Trait-anxiety | .78 | 51.05 (6.21) | 50.22 (4.85) | - | .44 |
| <i>Lesson 5</i> | | | | | |
| Performance | - | 609.86 (167.48) | 617.91 (152.28) | - | 1.00 |
| State-anxiety | .94 | 43.90 (6.62) | 48.50 (8.27) | .09 | .04 |
| <i>Deferred evaluation</i> | | | | | |
| Performance | - | 600.62 (154.66) | 681.32 (166.76) | - | .19 |
| State-anxiety | .93 | 49.57 (9.37) | 48.95 (8.96) | - | .78 |

Note: performance is expressed in centimeters; α = Cronbach's alpha; η^2 = eta-squared of Mann-Whitney U test with class factor for significant difference; p = p-value of Mann-Whitney U test with class factor for each variable to each time.

Time Effect for Each Class

Friedman's ANOVA including the first lesson, the last lesson and the deferred evaluation was conducted to observe whether there was a time effect on performance and state-anxiety in each assessment condition (see Figure 1). Regarding the level of performance, Friedman's ANOVA showed a trend for the formative class ($\chi^2 = 5.37$; $df = 2$; $p = .068$) and a significant effect with the summative class ($\chi^2 = 14.97$; $df = 2$; $p = .001$). The level of anxiety evolved significantly within the formative class ($\chi^2 = 7.79$; $df = 2$; $p = .020$), whereas the

summative class showed a non-significant evolution. Considering these results, Wilcoxon tests were applied between the first lesson and the last lesson and between the last lesson and the deferred evaluation (*i.e.*, four months later).

During the Learning Cycle

Formative Class. Regarding the level of performance, a significant increase was found between the first lesson and the last lesson ($z = 2.37; \eta^2 = .27; p = .018$). Concerning the level of state-anxiety, there was a significant decrease between the two lessons ($z = 3.14; \eta^2 = .47; p = .019$).

Summative Class. Regarding the level of performance, a significant increase was found between the first and the last lesson ($z = 2.61; \eta^2 = .31; p = .009$). There was no time effect for the level of state-anxiety during the learning cycle.

After the Learning Cycle

Formative Class. Regarding the level of performance, there was no time effect. However, the level of anxiety increased significantly between the last lesson and the deferred evaluation ($z = 2.72; \eta^2 = .35; p = .006$).

Summative Class. The level of performance increased significantly between the last lesson and the deferred evaluation ($z = 3.13; \eta^2 = .45; p = .002$). Concerning the level of state-anxiety, no time effect was observed.

Time*Class Effect for Each Variable

Mann-Whitney U tests were performed on progress in shot-put performance and state-anxiety between the first lesson and the last lesson, then between the last lesson and the deferred evaluation. These analyses were designed to observe differences in time effects between the two assessment conditions.

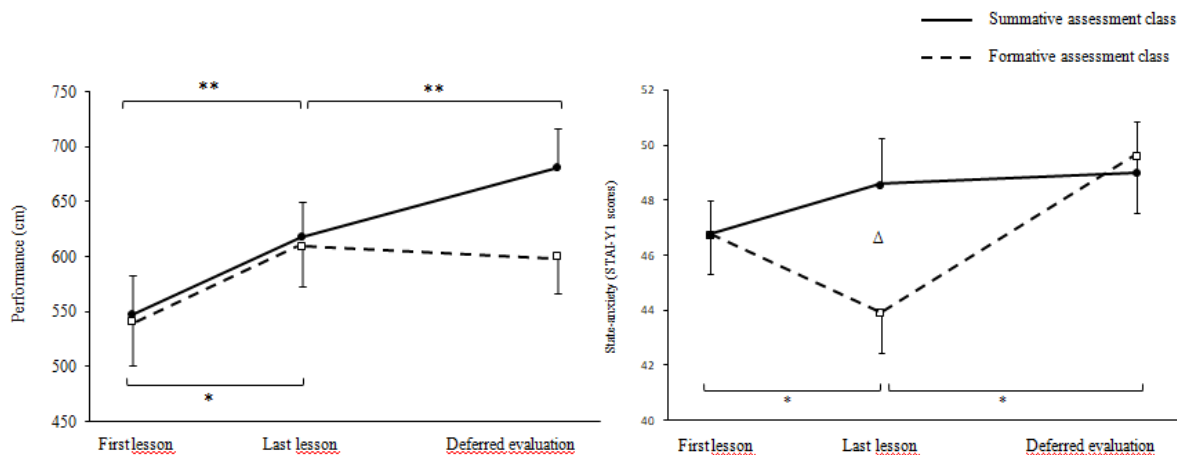
During the Learning Cycle

There was no time*class effect for the level of shot-put performance during the learning cycle. Concerning the level of state-anxiety, there was a significant difference in time effects between the two classes from the first to the last lesson in the learning cycle ($U = 311.50; z = 1.96; \eta^2 = .09; p = .050$).

After the Learning Cycle

There was a time*class effect for the level of shot-put performance between the last lesson in the learning cycle and the deferred evaluation ($U = 338.50; z = 2.61; \eta^2 = .11; p = .009$). Concerning the level of state-anxiety, there was a significant difference in time effects between the two classes from the last lesson in the learning cycle to the deferred evaluation ($U = 311.50; z = -2.25; \eta^2 = .12; p = .025$).

Figure 1. Time effect and class effect on performance and state-anxiety between the two classes in each lesson



Note : performance (expressed in centimeters), and state-anxiety (STAI Y1) for each class with the standard error of the mean (SEM); * = $p < .05$; ** = $p < .01$ for time effect on Wilcoxon test. $\Delta = p < .05$ for class effect on Mann-Whitney U test.

Discussion

The purpose of the present investigation was to compare the impact, between two classes, of assessment type (*i.e.*, formative vs summative) on the level of shot-put performance and state-anxiety. Analyses were conducted during the first and the last lessons of the learning cycle, then four months after the learning cycle. These analyses examined the class effect at each lesson, the time effect and the time*class effect. The data show that formative assessment seems to be associated with less state-anxiety than the summative assessment at the end of the learning cycle. This result confirms our first assumption that the type of assessment would

influence the level of state-anxiety in favor of formative assessment. Concerning the level of shot-put performance, both classes progressed similarly during the learning cycle. Most intriguingly, a difference in progress in shot-put performance was observed four months later in favor of the summative assessment class. This result invalidates our assumption and shows that students in the summative assessment class improved their performance, whereas the formative assessment class maintained its performance level.

During the Learning Cycle

Both classes significantly increased their level of shot-put performance between the first lesson and the last lesson in the learning cycle. There was no class effect or time*class effect on the level of shot-put performance. Contrary to our assumption, the formative assessment did not result in superior motor learning (*i.e.*, superior shot-put distance) compared to the summative assessment during the learning cycle. It should be noted that the students from both classes were novices at the beginning of the learning cycle and there were no differences in shot-put performance between students with throwing experience and those without throwing experience. It is possible that the improvement in performance among these novices may have appeared early and may have reached at a higher level without any deep transformation of their motor abilities such as coordination (Delignières et al., 2009). Thus, each repetition of the shot-put movement in each lesson represented an essential condition for motor learning. In addition and according to the differential learning approach (Henz & Schöllhorn, 2016) adopted in this study, some learning parameters may also have favored this improvement, such as learning through a series of methodical exercises (Gaulhofer & Streicher, 1924) and variability of practice (Schmidt, 1975; Schöllhorn et al., 2009). In the present study, the fact that there was no difference between the two classes at the end of the learning cycle suggests that the summative and formative assessment did not influence the shot-put motor learning. A recent study on PE (Kok et al., 2020) also reports no group effect on shot-put distance, under any experimental conditions (*i.e.*, three different types of video feedback).

In line with our assumption, state-anxiety decreased within the formative class and remained stable within the summative class between the first and the last lesson in the learning cycle. Thus, the formative class was less anxious than the summative class at the end of the learning cycle. These results are in agreement with those of Lentillon-Kaestner et al. (2018), who showed that classroom students without grades work with less stress. In the formative assessment class, students probably developed coping strategies or psychological adjustment capacities (Cosma et al., 2020; Naylor et al., 2001; Skaalvik, 2018). Moreover, these results are in line with (Biggs, 1998), who assumed that summative assessment was strongly linked to students' futures and therefore associated with strong emotions. However, it seems important to underline that the observed class effect on state-anxiety up to the last lesson was only due to the decrease in anxiety within the formative assessment class. Therefore, it would be more appropriate to suggest that formative assessment had a positive impact on emotional reaction, rather than assuming that summative assessment provoked state-anxiety.

After the Learning Cycle

There was no class effect on state-anxiety and shot-put performance four months later, but time and time*class effects on shot-put performance and state-anxiety were found between the end of the learning cycle and the deferred evaluation. A significant increase in performance was revealed within the summative assessment class, whereas the formative assessment class remained at the same level. In contrast, a significant increase in anxiety was revealed within the formative assessment class whereas the summative assessment class remained at the same level.

Four months later, the shot-put performance level of both classes was either maintained or increased. These benefits could be ascribed to the differential learning approach that was adopted as the learning method for this study (Henz & Schöllhorn, 2016). For example, benefits up to four weeks after the end of the intervention have also been found with the differential learning approach (Beckmann & Schöllhorn, 2006). In contrast, the same study found that performance decreased to return to its initial level as early as two weeks after the intervention when the learning method was based only on the exact prescriptions of a movement prototype and an intervention with numerous repetitions. Moreover, Scholz's studies on juggling may, to some extent, provide another explanation for these results. Scholz et al. (2009) reported an increase in gray matter after a 4-week period without training. In agreement with these data, we could speculate that structural cerebral and/or functional changes due to learning practice, which occurred after the learning cycle, are plausible explanations for the shot-put performance being maintained or increased four months later.

In addition, it should be noted that the difference in progress in shot-put performance between the two classes could be multifactorial. Among these factors, state-anxiety experiences during the learning cycle may have influenced the progress observed in shot-put performance four months later. Indeed, students in the formative assessment class were less anxious during the learning cycle than students in the summative assessment class. Under these conditions, the encoding stage and the quality of retrieval could be better for the summative class. Indeed, thanks to amygdala stimulation, emotional memories are more reliable and robust than non-emotional memories (*e.g.*, Kensinger & Kark, 2018). Therefore, the summative assessment class could have memorized shot-put motor learning better during the learning cycle thanks to its emotional

characteristics. Furthermore, the summative assessment class may have perceived the grade as the sole objective of the learning cycle, generating greater external motivation (Barnett, 2007) than the formative assessment class. In agreement with Lentillon-Kaestner et al. (2018), who found that students and PE teachers rated the grade as having a positive impact on engagement and achievements, we assume that this teaching style had a positive effect on performance for the summative assessment class.

Lastly, the increased anxiety associated with the deferred evaluation could be linked to the difficulty of performing a task that was learned four months ago. As mentioned, it may be possible that the encoding stage and the quality of retrieval were lower for the formative assessment class. These results suggest that the benefits of formative assessment incorporating strategies are not sustainable. This finding appears to be inconsistent with most studies, which report that formative assessment brings many benefits (e.g., Black & Wiliam, 1998; Harlen et al., 2002).

Limitations and Future Research

Several limitations need to be considered when interpreting the outcomes of this study. The STAI (Spielberger, 1983) is a highly developed and validated tool, but not specifically adapted to PE. Future research should validate a French version of the “Physical Education State-Anxiety Scale” (PESAS, Barkoukis et al., 2008) in order to use it in ecological studies. Another limitation was that the presence / absence of grades in each lesson was only one factor among others, affecting the educational climate. For example, the present study only considered the type of assessment without testing teaching methods such as the global motivational climate or the quality of feedback provided to the students. Future research should complement the measures with motivational climate through observation (MMCOS; Smith et al., 2015) or motivational questionnaires. Future studies analyzing the effects of assessment on motor performance should include age and sex variables because these variables may influence the motor coordination (Benjumea et al., 2015). Other dispositional variables as personality resources (e.g., self-efficacy, self-control, optimism and grit) could be considered for further researches analyzing sport anxiety (Olefir, 2018).

Conclusions

This pilot study, which aimed to provide an insight into grades in PE, has produced modest but interesting results which deserve to be taken further in future. Results showed that both classes had improved their shot-put performance during the learning cycle. However, the formative assessment class felt less anxiety than the summative assessment class at the end of the learning cycle. Four months later, the summative assessment class's performance had improved, whereas the formative assessment class's performance had remained at the same level. In contrast, the summative assessment class's anxiety had remained at the same level, whereas the formative assessment class's anxiety had improved. This study adds to the debate around “formative good /summative bad” dichotomy (Lau, 2016) because it shows the positive impact of summative assessment on motor learning and unstable benefits of formative assessment on emotional reactions in PE. However, based on ethical and pedagogical considerations and given that a positive sport experience during childhood favors an active lifestyle during adulthood (Doolittle & Rukavina, 2014; Telama et al., 2005), can we really reduce the students' positive experience down to better learning? Since the formative assessment class remained at the same level of performance four months later, with students feeling less anxious at the end of the learning cycle, we suggest that PE lessons should promote formative assessment. It would also be interesting to consider that formative and summative assessment need to be implemented in harmony and not in opposition to each other (Lau, 2016). Lastly, future studies aiming for a better understanding of the influence of the type of assessment on students' motor learning and emotional reaction should consider attentional and motivational mechanisms (OPTIMAL theory; Wulf & Lewthwaite, 2016). Indeed, motivation influences students' affective experiences such as anxiety (Hogue et al., 2017). Anxiety has an effect on attentional processes (Eysenck et al., 2007; Payne et al., 2019) that is central to sports performance (Furley & Wood, 2016; Nideffer, 1976) and motor learning (Wulf, 2013).

Conflicts of Interest Statement and Funding Details: None

References

- Alcalá, D. H., Picos, A. P., & Pastor, V. L. (2019). The impact of formative and shared or co-assessment on the acquisition of transversal competences in higher education. *Assessment & Evaluation in Higher Education*, 44(6), 933–945. <https://doi.org/10.1080/02602938.2018.1530341>
- Barkoukis, V., Tsozbatzoudis, H., & Grouios, G. (2008). Construct Validity of the Physical Education State Anxiety Scale: A Multitrait—Multimethod Approach. *Perceptual and Motor Skills*, 107(3), 651–662. <https://doi.org/10.2466/pms.107.3.651-662>
- Barnett, R. (2007). Assessment in higher education: An impossible mission? In *Rethinking Assessment in Higher Education* (pp. 39–50). Routledge. <https://doi.org/10.4324/9780203964309-10>

- Beckmann, H., & Schöllhorn, W. I. (2006). Differenzielles Lernen im Kugelstoßen [Differential learning in shot put]. *Leistungssport*, 36(4), 44–50.
- Benjumea, J. M. C., Afonso, J. R., Hurtado, J. M. R., & Truan, J. C. F. (2015). Assessment of Motor Coordination in Students Aged 6 to 11 Years. *Journal of Physical Education and Sport*, 15(4), 765.
- Benmansour, N. (1999). Motivational orientations, self-efficacy, anxiety and strategy use in learning high school mathematics in Morocco. *Mediterranean Journal of Educational Studies*, 4(1), 1–15.
- Biggs, J. (1998). Assessment and classroom learning: A role for summative assessment? *Assessment in Education: Principles, Policy & Practice*, 5(1), 103–110. <https://doi.org/10.1080/0969595980050106>
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Broadfoot, P. (2002). Beware the Consequences of Assessment! *Assessment in Education: Principles, Policy & Practice*, 9(3), 285–288. <https://doi.org/10.1080/0969594022000027645>
- Bruchon-Schweitzer, M., & Paulhan, I. (1993). French adaptation of Spielberger's Trait-Etat (Form Y) anxiety inventory [Adaptation francophone de l'inventaire d'anxiété Trait-Etat (Forme Y) de Spielberger]. *Editions Du Centre Psychologie Appliquée*.
- Carless, D., & Lam, R. (2014). The examined life: Perspectives of lower primary school students in Hong Kong. *Education 3-13*, 42(3), 313–329. <https://doi.org/10.1080/03004279.2012.689988>
- Cauley, K. M., & McMillan, J. H. (2010). Formative Assessment Techniques to Support Student Motivation and Achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 83(1), 1–6. <https://doi.org/10.1080/00098650903267784>
- Chee-Cheong, C. (1999). Public Examinations in Hong Kong. *Assessment in Education: Principles, Policy & Practice*, 6(3), 405–417. <https://doi.org/10.1080/09695949992829>
- Cosma, G., Chiracu, A., Stepan, R., Cosma, A., Nanu, C., & Paunescu, C. (2020). Impact of coping strategies on sport performance. *Journal of Physical Education and Sport*, 20(3), 1380–1385. <https://doi.org/10.7752/jpes.2020.03190>
- Delignières, D., Teulier, C., & Nourrit, D. (2009). Learning complex motor skills: From spontaneous coordination to expert coordination [L'apprentissage des habiletés motrices complexes: Des coordinations spontanées à la coordination experte]. *Bulletin de psychologie*, 502(4), 327–334.
- Doolittle, S. A., & Rukavina, P. B. (2014). Case Study of an Institutionalized Urban Comprehensive School Physical Activity Program. *Journal of Teaching in Physical Education*, 33(4), 528–557. <https://doi.org/10.1123/jtpe.2014-0073>
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, 7(2), 336–353. <https://doi.org/10.1037/1528-3542.7.2.336>
- Furley, P., & Wood, G. (2016). Working Memory, Attentional Control, and Expertise in Sports: A Review of Current Literature and Directions for Future Research. *Journal of Applied Research in Memory and Cognition*, 5(4), 415–425. <https://doi.org/10.1016/j.jarmac.2016.05.001>
- Gaulhofer, K., & Streicher, M. (1924). *Basics of Austrian school gymnastics* [Grundzüge des österreichischen Schulturnens]. Fundamentals of Austrian school gymnastics.
- Gentile, A. M. (1972). A Working Model of Skill Acquisition with Application to Teaching. *Quest*, 17(1), 3–23. <https://doi.org/10.1080/00336297.1972.10519717>
- Gipps, C. V. (2002). *Beyond testing: Towards a theory of educational assessment*. Routledge.
- Harlen, W., Crick, R. D., Broadfoot, P., Daugherty, R., Gardner, J., James, M., & Stobart, G. (2002). A Systematic Review of the Impact of Summative Assessment and Tests on Students' Motivation for Learning. *EPPI-Centre, University of London*. <http://dspace.stir.ac.uk/handle/1893/19607>
- Henz, D., & Schöllhorn, W. I. (2016). Differential Training Facilitates Early Consolidation in Motor Learning. *Frontiers in Behavioral Neuroscience*, 10, 199. <https://doi.org/10.3389/fnbeh.2016.00199>
- Hogue, C. M., Fry, M. D., & Fry, A. C. (2017). The differential impact of motivational climate on adolescents' psychological and physiological stress responses. *Psychology of Sport and Exercise*, 30, 118–127. <https://doi.org/10.1016/j.psychsport.2017.02.004>
- Hubiche, J.-L., & Pradet, M. (1993). *Understanding athletics* [Comprendre l'athlétisme]. INSEP.
- Ianovici, E., & Weissbluth, E. (2016). Effects of Learning Strategies, Styles and Skill Level on Closed and Semi-Open Motor Skills Acquisition. *Journal of Physical Education and Sport*, 16(4), 1169.
- Kensinger, E. A., & Kark, S. M. (2018). Emotion and Memory. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (pp. 1–26). American Cancer Society. <https://doi.org/10.1002/9781119170174.epcn101>
- Knapp, B. (1963). *Skill in sport: The attainment of proficiency*. Routledge.
- Knight, P. T. (2002). Summative Assessment in Higher Education: Practices in disarray. *Studies in Higher Education*, 27(3), 275–286. <https://doi.org/10.1080/03075070220000662>
- Kok, M., Komen, A., Capelleveen, L. van, & Kamp, J. van der. (2020). The effects of self-controlled video feedback on motor learning and self-efficacy in a Physical Education setting: An exploratory study on the

-
- shot-put. *Physical Education and Sport Pedagogy*, 25(1), 49–66. <https://doi.org/10.1080/17408989.2019.1688773>
- Krijgsman, C., Vansteenkiste, M., van Tartwijk, J., Maes, J., Borghouts, L., Cardon, G., Mainhard, T., & Haerens, L. (2017). Performance grading and motivational functioning and fear in physical education: A self-determination theory perspective. *Learning and Individual Differences*, 55, 202–211. <https://doi.org/10.1016/j.lindif.2017.03.017>
- Kwon, S. K., Lee, M., & Shin, D. (2017). Educational assessment in the Republic of Korea: Lights and shadows of high-stake exam-based education system. *Assessment in Education: Principles, Policy & Practice*, 24(1), 60–77. <https://doi.org/10.1080/0969594X.2015.1074540>
- Lau, A. M. S. (2016). ‘Formative good, summative bad?’ – A review of the dichotomy in assessment literature. *Journal of Further and Higher Education*, 40(4), 509–525. <https://doi.org/10.1080/0309877X.2014.984600>
- Law, W.-W. (2007). Legislation and educational change: The struggle for social justice and quality in China’s compulsory schooling. *Education and the Law*, 19(3–4), 177–199. <https://doi.org/10.1080/09539960701751493>
- Leighton, J. P. (2019). Students’ Interpretation of Formative Assessment Feedback: Three Claims for Why We Know So Little About Something So Important. *Journal of Educational Measurement*, 56(4), 793–814. <https://doi.org/10.1111/jedm.12237>
- Lentillon-Kaestner, V., Deriaz, D., Voisard, N., & Allain, M. (2018). *Grade in physical education? Implications for teaching and students [Noter en éducation physique? Incidences sur l’enseignement et les élèves]*. EME éditions. <https://orfee.hepl.ch/handle/20.500.12162/159>
- Linthorne, N. P. (2001). Optimum release angle in the shot put. *Journal of Sports Sciences*, 19(5), 359–372. <https://doi.org/10.1080/02640410152006135>
- Martos-Garcia, D., Usabiaga, O., & Valencia-Peris, A. (2017). Students’ Perception on Formative and Shared Assessment: Connecting Two Universities through the Blogosphere. *Journal of New Approaches in Educational Research*, 6(1), 64. <https://doi.org/10.7821/naer.2017.1.194>
- McDonald, A. S. (2001). The Prevalence and Effects of Test Anxiety in School Children. *Educational Psychology*, 21(1), 89–101. <https://doi.org/10.1080/01443410020019867>
- Naylor, P., Cowie, H., & del Rey, R. (2001). Coping strategies of secondary school children in response to being bullied. *Child Psychology & Psychiatry Review*, 6(3), 114–120. <https://doi.org/10.1017/S1360641701002647>
- Nideffer, R. M. (1976). Test of attentional and interpersonal style. *Journal of Personality and Social Psychology*, 34(3), 394–404. <https://doi.org/10.1037/0022-3514.34.3.394>
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.
- Olefir, V. (2018). Personality Resources as a Mediator of the Relationship between Antecedents of Stress and Pre-Competitive Anxiety. *Journal of Physical Education and Sport*, 18(4), 2230.
- Payne, K. L., Wilson, M. R., & Vine, S. J. (2019). A systematic review of the anxiety-attention relationship in far-aiming skills. *International Review of Sport and Exercise Psychology*, 12(1), 325–355. <https://doi.org/10.1080/1750984X.2018.1499796>
- Pollard, A., & Triggs, P. (2001). *What Pupils Say: Changing Policy and Practice in Primary Education* (A & C Black). A&C Black.
- Putwain, D. W., & Best, N. (2011). Fear appeals in the primary classroom: Effects on test anxiety and test grade. *Learning and Individual Differences*, 21(5), 580–584. <https://doi.org/10.1016/j.lindif.2011.07.007>
- Raiola, G. (2017). Motor Learning and Teaching Method. *Journal of Physical Education and Sport*, 17, 2239.
- Reay, D., & Wiliam, D. (1999). ‘I’ ll be a nothing’: Structure, agency and the construction of identity through assessment. *British Educational Research Journal*, 25(3), 343–354. <https://doi.org/10.1080/0141192990250305>
- Roderick, M., & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational Evaluation and Policy Analysis*, 23(3), 197–227. <https://doi.org/10.3102/01623737023003197>
- Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, 82(4), 225–260. <https://doi.org/10.1037/h0076770>
- Schöllhorn, W. I., Mayer-Kress, G., Newell, K. M., & Michelbrink, M. (2009). Time scales of adaptive behavior and motor learning in the presence of stochastic perturbations. *Human Movement Science*, 28(3), 319–333. <https://doi.org/10.1016/j.humov.2008.10.005>
- Scholz, J., Klein, M. C., Behrens, T. E. J., & Johansen-Berg, H. (2009). Training induces changes in white-matter architecture. *Nature Neuroscience*, 12(11), 1370–1371. <https://doi.org/10.1038/nn.2412>
- Skaalvik, E. M. (2018). Mathematics anxiety and coping strategies among middle school students: Relations with students’ achievement goal orientations and level of performance. *Social Psychology of Education*, 21(3), 709–723. <https://doi.org/10.1007/s11218-018-9433-2>

- Smith, N., Tessier, D., Tzioumakis, Y., Quested, E., Appleton, P., Sarrazin, P., Papaioannou, A., & Duda, J. L. (2015). Development and validation of the Multidimensional Motivational Climate Observation System. *Journal of Sport & Exercise Psychology*, 37(1), 4–22. <https://doi.org/10.1123/jsep.2014-0059>
- Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory STAI (form Y) ('self- 839 evaluation questionnaire')*.
- Telama, R., Yang, X., Viikari, J., Välimäki, I., Wanne, O., & Raitakari, O. (2005). Physical activity from childhood to adulthood: A 21-year tracking study. *American Journal of Preventive Medicine*, 28(3), 267–273. <https://doi.org/10.1016/j.amepre.2004.12.003>
- Volante, L., & Beckett, D. (2011). Formative assessment and the contemporary classroom: Synergies and tensions between research and practice. *Canadian Journal of Education / Revue Canadienne de l'éducation*, 34(2), 239–255. JSTOR.
- Wulf, G. (2013). Attentional focus and motor learning: A review of 15 years. *International Review of Sport and Exercise Psychology*, 6(1), 77–104. <https://doi.org/10.1080/1750984X.2012.723728>
- Wulf, G., & Lewthwaite, R. (2016). Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychonomic Bulletin & Review*, 23(5), 1382–1414. <https://doi.org/10.3758/s13423-015-0999-9>
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, 45(4), 477–501. <https://doi.org/10.1023/A:1023967026413>